Global Headquarters: 5 Speen Street Framingham, MA 01701 USA P.508.872.8200 F.508.935.4015 www.idc.com

WHITE PAPER

Intel's Enterprise Processor Plans: Positioning the Xeon Processor and the Itanium Processor

Sponsored by: Intel Corporation

Vernon Turner

April 2004

Executive Summary

Intel will ship an enhanced version of its Xeon processor in mid-2004. In addition to offering full native 32-bit performance, expanded on-chip cache capacity, new power management features, and faster front-side bus (FSB), this new Xeon processor will include Intel Extended Memory 64 Technology (Intel EM64T) to extend the physical and virtual address spaces for the Xeon family of processors.

Intel will offer two 64-bit processor product lines: Xeon processors with 64-bit extensions and Itanium processors. IDC believes that these two product lines will be differentiated and will coexist in the marketplace. While the processors both contain 64-bit registers and support larger address spaces, they differ in other areas and will be used to support different workloads.

The new Xeon processor will be especially useful for the large number of 32-bit applications that can be run with full native performance. ISVs and users with custom software can recompile their code to access more memory. At the same time, the Itanium processor offers a larger address space; larger caches; reliability, availability, and serviceability (RAS) features; and higher performance from a new processor architecture: explicitly parallel instruction computing (EPIC).

Xeon and Itanium processors will be used to build smaller scale-out servers, and Itanium processors will be used to build larger scale-up servers. A scale-out architecture provides more capacity by adding a greater number of smaller servers, while a scale-up architecture provides more capability by adding a smaller number of larger servers. Workloads that do not share data extensively and can be decomposed and executed in parallel as independent subprocesses are best supported by a scale-out architecture. Workloads that do share data extensively but are inherently multitasked or can be decomposed into dependent parallel tasks are best supported by a scale-up architecture that calls for a more powerful server. For many applications and workloads, both scale-up and scale-out solutions will be technically viable and price/performance ratios will play the final role in IT decision making.

IDC believes that Intel's introduction of the Xeon processor with 64-bit extensions provides OEMs and their customers with an important new choice in a highly competitive market. The ability of servers to host both legacy 32-bit applications at native performance and new 64-bit applications will ease the path forward for IT organizations and their ISVs and accelerate migration to a 64-bit industry-standard platform.

Introduction

Customers continue to demand servers and workstations with improved performance and solid reliability at the lowest possible cost. Among commercial workloads, however, some applications are as computationally demanding as traditional high-performance applications. Such highly demanding workloads include data mining, business intelligence, and financial modeling. Video and other visual media are increasingly prevalent and present performance challenges not only to support real-time image and video algorithms but also to access large files. These new and demanding workloads typically share the following characteristics:

Customers continue to demand servers and workstations with improved performance and solid reliability at the lowest possible cost.

- Supporting most server workloads increasingly means retrieving, processing, and storing large data sets; supporting multiple users; and providing reliable service to the many applications that have become mission critical in the enterprise.
- Supporting some workstation workloads means executing high-performance integer operations and algorithms that demand large amounts of working memory.

At the same time, many commercial workloads are comfortably supported by servers with current 32-bit processors. Results from IDC's 4Q03 Worldwide Quarterly Server Tracker indicate that approximately 85% of servers sold for prices below $6,000 and contained 1GB of memory or less, which is far less than the 32-bit limit of 4GB. Typical examples of 32-bit workloads are infrastructure servers that carry email and Web services workloads. These applications do not demand a large amount of addressable memory. Furthermore, these applications are written with programming tools designed for 32-bit processors.

[M]any commercial workloads are comfortably supported by servers with current 32-bit processors.

During recent years, IT system architects have categorized datacenter server resources as 32-bit versus 64-bit and matched available software to the appropriate class of server. In the future, however, these system architects will be able to set aside the 32-bit/64-bit distinction in favor of other platform characteristics such as performance, price performance, RAS, and scale when they match workloads to servers. And, industry-standard servers will support a 64-bit address space.

Intel Xeon Processor Enhancements

Cognizant of these trends, Intel has announced an evolution to its Xeon processor. 64-bit addressing has been added to a number of other processor- and platform-level innovations that are unrelated to 64-bit addressing. The Xeon processor with 64-bit extensions is backward compatible with the IA-32 instruction set and with software written for 32-bit Xeon processors. This next-generation Xeon processor has full native 32-bit performance, new instructions, faster FSB, new power management features, greater on-chip cache sizes, and Intel EM64T, which includes 64-bit registers and larger physical address space. Table 1 summarizes these innovations.

Intel has announced an evolution to its Xeon processor. 64-bit addressing has been added to a number of other processor- and platform-level innovations that are unrelated to 64-bit addressing.

TABLE 1

Processor and Platform Enhancements for the Intel Xeon Processor

Enhancement	Function	Feature	Description
Processor	64-bit processing	Extended Memory 64 Technology (EM64T)	EM64T includes extended memory addressability (i.e., 64-bit pointers and registers), eight additional registers for Streaming SIMD [Single Instruction, Multiple Data] Extensions (SSE) and eight additional general-purpose registers, and double-precision integer support. A new, extended IA-32 mode will run both 32-bit and 64-bit applications side by side under a 64-bit operating system.
Processor	Bus bandwidth	800MHz front-side bus (FSB)	The FSB provides the processor with more bandwidth and faster access to memory.
Processor	Instruction set	SSE3 instructions	SIMD programming processes more data with fewer instructions, thus improving performance. SSE3 enables accelerated processing of complex arithmetic and video decoding.
Processor	Power management	Demand-based switching (DBS)	DBS minimizes energy consumption and heat by changing processor performance or power state (i.e., voltage or clock frequency) in response to processor utilization.
Platform	Input/output (I/O)	PCI Express	PCI Express is a next-generation, direct-attach serial I/O for lower latency and higher-bandwidth I/O and board-level interconnect. PCI Express will also replace the accelerated graphics port (AGP) interconnect for graphics.
Platform	Memory	DDR2 memory	DDR2 is faster and higher-capacity memory with lower power requirements.

Source: Intel, 2004

64-Bit Processors

64-bit processors have registers and arithmetic logic units capable of manipulating 64 bits of data at each processing step, and they are also capable of addressing larger amounts of memory. A processor with a 64-bit address space allows programs to directly access 16 exabytes (EB) of memory, a massive increase from the maximum 4GB of memory addressed by 32-bit processors. Although 4GB is a substantial amount of data, some of today's business applications address even larger data sets. For example, enterprise resource planning (ERP) and data mining (i.e., large database analysis) applications often exceed the 4GB limit. In these cases, a 32-bit processor will require additional clock cycles and cumbersome programming techniques to access memory beyond 4GB, thus negatively affecting performance.

A processor's address space is the range of values that the processor uses to access data stored in memory. A physical address space maps values directly to memory locations in hardware devices. A virtual address space simulates a memory resource larger than actual physical memory to be used by application programs. At runtime, hardware and software maintain a mapping between virtual and physical memory and a swapping of data in physical memory locations.

The Xeon processor with 64-bit extensions supports 48 bits of virtual memory and up to 40 bits of physical memory. Therefore, programmers writing code for servers with Xeon processors have access to 256TB of virtual memory, and those servers can address up to 1TB of physical memory. In contrast, the Itanium processor supports 64-bit virtual memory and 50-bit physical memory, which corresponds to 16EB of virtual memory and 1 petabyte (PB) of physical memory.

Larger registers and address spaces do not inherently improve a processor's performance. Performance will not increase when applications that require less memory move from a 32-bit processor to a 64-bit processor with the same instruction set. When a 32-bit application is run on a 64-bit processor, software still limits memory access to 4GB and performance might not improve. Recompilation will be required to take advantage of memory in excess of 4GB. An application recompiled to run in a 64-bit environment will be able to access significantly more memory, and performance might improve if the application was previously memory constrained.

Workloads Best Suited for Servers with Xeon Processors

Intel identifies three key workloads for which the Xeon processor with 64-bit extensions is well suited:

- The primary workloads for the enhanced Xeon processor will continue to be legacy 32-bit enterprise applications. These workloads can profit from the new Xeon processor's greater processing speed and larger caches. Because no recompilation is needed to run 32-bit applications on servers using Xeon processors with 64-bit extensions, IT planners will be able to prepare for a gradual migration to 64-bit computing with a uniform inventory of 64-bit servers. Note that a new 64-bit operating system and device drivers will be needed to access the 64-bit capabilities of the enhanced Xeon processors. 32-bit applications will need to be recertified for the 64-bit operating environment even if the software remains 32-bit.

- The new Xeon processor will support integer-intensive high-performance technical computing workloads. Integer performance for the new Xeon processor is approximately equal to the integer processing performance of the Intel Itanium 2 processor. Thus, Intel expects servers based on Xeon processors with 64-bit extensions to be used as nodes that scale out to provide greater capacity clusters. Intel also expects the new Xeon processor to be used in workstations that support engineering design applications where a combination of high-performance 32-bit applications will be run concurrently with a small number of large data set 64-bit applications.

- Some infrastructure services will profit from the Xeon processor's EM64T and its other enhancements. Web and email services are examples of infrastructure workloads that are growing rapidly. While not as demanding as other enterprise applications, these workloads might profit from access to a full 4GB address space that is not shared with system software. When packaged software providers migrate business application products to the new Xeon processor with 64-bit extensions, then IT planners can scale up to servers with increased capability or scale out with additional servers to add capacity.

Scaling Up and Scaling Out: Capability and Capacity Workloads

When IDC analyst Christopher Willard described the supercomputer industry in September 1999, he noted that clusters of commercial-class computers could address some computationally intense workloads while large, monolithic supercomputers could meet the demands of other workloads. Willard distinguished between *capacity* systems and *capability* systems, respectively, and server analysts at IDC have used this distinction ever since.

Large problems that decompose into small, relatively independent tasks can be processed with a cluster of computing machines, and the processing capacity can be accelerated by adding more nodes to the cluster. Meanwhile, other problems that cannot be decomposed require a system with greater capability (i.e., more powerful processing and larger addressable memory).

For many years, IDC's distinction between capability and capacity has been a helpful framework for understanding technical and scientific workloads along with the high-performance technical computing environments designed to carry those workloads.

More recently, IT architects have begun to talk about a key design decision in the datacenter: Should computing resources be scaled up by deploying a smaller number of larger servers, or should computing resources be scaled out by deploying a larger number of smaller servers? Scaling up means adding a few capable servers, and scaling out means adding more capacity servers.

IDC derives three important lessons from the capability/capacity distinction:

- Workload characteristics determine when a capability approach is necessary. Scaling up a system with more powerful servers will be required when the workload cannot be decomposed into smaller, independent subprocesses.
- Many workloads thought initially to require a capability system have been successfully decomposed, and, as a result, most of today's high-performance computing (HPC) systems (i.e., supercomputers) can best be described as capacity systems.
- Most commercial workloads can be supported by either a capability or a capacity design — that is, by scaling up or by scaling out. As a result, price/performance characteristics of the alternative solutions weigh heavily in the decision making.

Xeon and Itanium Processors: Three Key Differences

- Intel Xeon processors with 64-bit extensions will provide full native 32-bit performance for the large number of 32-bit applications while allowing code that is recompiled for 64-bit processing to access more memory. Intel Itanium processors will run IA-32 software, but not with full native 32-bit Xeon processor performance. Itanium processors not only provide a larger address space but also bring to market larger caches and new performance-enhancing technologies, such as the EPIC architecture.

- Integer processing performance is approximately equal for Xeon processors with 64-bit extensions and Itanium processors, while floating-point processing is significantly faster on Itanium processors. Transaction and enterprise application performance is also faster on Itanium processors than on Xeon processors. These performance differences mean that Itanium processors are better suited for database and high-end enterprise applications and for many scientific and analytic applications.

- Xeon processors with 64-bit extensions and Itanium processors are both engineered with on-chip reliability and scalability enablers, including memory spares, chipkill memory, and error correction codes (ECC). The Itanium processor family employs additional on-chip reliability techniques — such as the ability to contain bad data — that favor its use for mission-critical applications. In addition, servers are more scalable when composed of Itanium processors rather than Xeon processors with 64-bit extensions due to the Itanium processor's larger caches and higher bus bandwidth. Table 2 contrasts platforms using the new Xeon processor with those using the Itanium processor.

TABLE 2

Contrasting Platforms Using Xeon Processors and Itanium Processors

Attribute	Platform with Xeon Processors	Platform with Itanium Processors
Performance	Best for workgroup, workstation, Web server, and IA-32 applications	Best for largest enterprise database and technical computing workloads
Reliability, availability	Reliable data integrity	Data integrity and high availability to replace RISC and mainframe processors at a fraction of the cost
Scalability	Scalable to 4x, 8x, and 16x processors, and beyond, but with decreasing performance improvement; 8x platform with Xeon processors is currently about 15% higher performance than 4x platform with Itanium processors	Scalable with increasingly better performance over platforms using the Xeon processor as processor counts increase; 4x servers with Itanium processors outperform 4x servers with Xeon processors by about 35%, whereas 32x servers with Itanium processors outperform 32x servers with Xeon processors by about 140%
Physical addressing	Supports up to 1TB of addressable memory	Supports up to 1PB (1,000TB) of addressable memory; memory capacity for largest symmetric multiprocessor (SMP) platforms and enterprise data sets
Platform	Balanced platform bandwidth and high performance for delivering leading-edge technologies	High bus bandwidths, scalability, and platform longevity for end-user investment protection
Solutions	Enterprise infrastructure, Web, mail, customer relationship management, and supply chain management (logistics); high-performance capacity applications; digital content creation; mechanical computer-aided design, electronic design automation	Enterprise database, online transaction processing, business intelligence, supply chain management (planning), enterprise resource planning, high-performance capability applications, computer-assisted engineering

Source: Intel, 2004

Xeon and Itanium Processors: Workload Analysis

IT planners should weigh the relative strengths of Xeon with 64-bit extensions and Itanium processors when making deployment decisions. As Figure 1 shows, Intel generally expects both Xeon and Itanium processors to play a role in all enterprise architecture tiers and in HPC platforms. Key differences in performance, scalability, and reliability lead to different optimal workloads for workstations and servers built with Xeon versus Itanium processors:

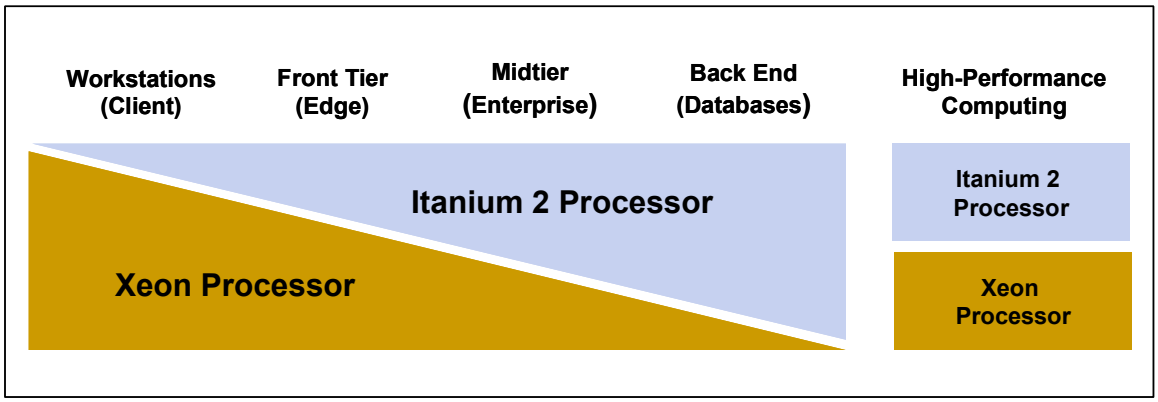
IT planners should weigh the relative strengths of Xeon with 64-bit extensions and Itanium processors when making deployment decisions.

- The Xeon processor with 64-bit extensions is best suited for infrastructure, workgroup, workstation, Web, and IA-32 applications in the datacenter. With full-performance IA-32 compatibility and 64-bit address space combined with Moore's law–based improvements in performance, Intel's new Xeon processor will handle the increasing requirements for these enterprise workloads.

- The Itanium processor is best suited for enterprise-scale database and transaction processing workloads and for technical computing applications. With its EPIC architecture, larger caches, higher bus bandwidth, and greater scalability, the Itanium processor will often be deployed in large symmetric multiprocessor (SMP) servers to provide greater performance for these workloads than the new Xeon processor. Scaling advantages for SMP servers using the Itanium processor versus the new Xeon processor increase sharply when moving to 8-way and higher SMP configurations.

FIGURE 1

Intel Platforms Span the Enterprise



Source: Intel, 2004

Both Intel processors offer industry-standard architectures that are well supported by system suppliers and software vendors. The new Xeon processor will initially support 64-bit Linux operating environments; support for the 64-bit Microsoft Windows operating system is expected in the second half of 2004. The Itanium processor already supports 64-bit Windows, Linux, and Unix operating systems. Scale-out deployment of multiple servers configured as clusters or grids of pooled computing resources will be common — an emerging architecture that IDC calls the *utility computing model*.

The Business Value Impact of Xeon Processors with 64-Bit Extensions

Intel's new Xeon processor addresses a broader range of computing requirements and will be used to support both commercial and HPC workloads. Pricing for Xeon and Itanium processors will evolve during the next few years. The Xeon processor with 64-bit extensions will lead to a steady increase in 64-bit applications.

Pricing Strategy for Xeon and Itanium Platforms

During the next three years, Intel expects cost convergence for platforms (i.e., server systems) using Xeon and Itanium processors, with Itanium processor–based platforms dropping in cost and achieving parity by 2007. Converged pricing means that the end user's decision to select and purchase a server using Xeon or Itanium processors will be based more on workload characteristics and OEM differentiation of products and less on platform cost differential.

To make possible converging platform prices, Intel is driving toward greater commonality in the Itanium and Xeon processor platforms. Common chipset and power components will be needed to enable OEMs to build servers that fit within a single pricing structure.

The Intel Architecture Ecosystem

Customers who integrate servers and workstations with Xeon processors with 64-bit extensions and Itanium processors will avail themselves not only of Intel's processor technology but also of a large ecosystem of product and service providers. Major system suppliers worldwide will ship servers with both processors, software suppliers will port to both server systems, and integrators will deploy these systems.

The process takes time, however; software suppliers typically spend up to a year porting (i.e., migrating, optimizing, and certifying) versions of their products for new processors after hardware and operating systems are available in the marketplace. IDC believes that enough of the software ecosystem for Xeon processors with 64-bit extensions will be in place in 2005 so that end users will begin developing and deploying enterprise solutions. More software tools and compilers as well as the thousands of packaged software solutions will be certified for 64-bit execution on Xeon processors in 2006 and 2007.

Intel expects system suppliers to draw from the Itanium processor family when building large SMP servers and mission-critical, highly available systems. Volume platforms and high-performance workstations will primarily use the Xeon processor with 64-bit extensions.

IDC Analysis

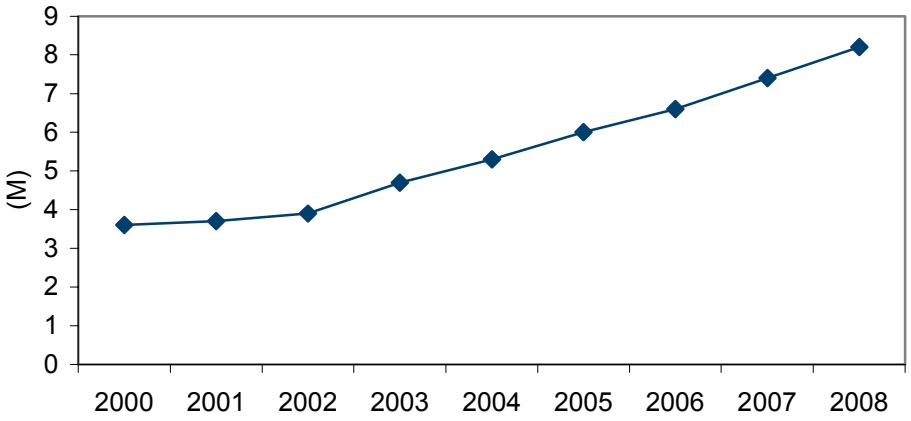
The market for workstations and servers using processors that support the IA-32 architecture is large and attractive to processor manufacturers and system vendors alike. To date, Intel and its competitors have designed processors and platforms to address specific segments of this market. Low-power processors for laptops and Centrino-enabled platforms to enable wireless networking are two examples of these specialty markets within the broader world of IA-32 processors.

The Xeon processor with 64-bit extensions is Intel's response to both direct competition and market opportunity. While Intel intends for servers based on the Itanium processor to be the most powerful and reliable in the marketplace, IT planners have long recognized that 32-bit IA-32 platforms are cost-effective and powerful enough to support many datacenter needs. By launching the Xeon processor with 64-bit extensions, Intel is endorsing the long-term viability of the IA-32 architecture. By converging price points for Xeon and Itanium platforms, Intel is communicating to its customers and OEMs its intent to supply flexible and complementary technology offerings.

The market for processors that support the x86 instruction set will continue to expand. IDC's forecast for growth in the x86 market, which is the market segment that includes Xeon processors with 64-bit extensions, is shown in Figure 2. IDC expects shipments of servers with x86 processors to rise to 8.2 million units in 2008 from 5.3 million units in 2004.

FIGURE 2

Worldwide x86 Server Processor Unit Shipments, 2000–2008



Source: IDC's 4Q03 Worldwide Quarterly Server Tracker

Opportunities

By bringing the Xeon processor with 64-bit extensions to market, Intel is reaching out to defined segments of the workstation and server markets with a processor tailored to their needs. The portfolio of processor offerings from Intel would be incomplete without the Itanium processor and the new Xeon processor offering.

The Xeon processor offers a platform for an orderly transition from 32-bit to 64-bit application software. Investment in servers using the Xeon processor will yield an environment that can run legacy applications along with workloads using software that has been recompiled to utilize the Xeon processor's new capabilities. Software revised and recompiled for 64-bit Xeon processors will be an important step closer to running on platforms using Itanium processors.

In the next generations of Xeon and Itanium processors, 64-bit processing equality should help customers to focus on workload requirements and no longer be confused or misdirected by the width of the processor's registers. Servers using Xeon and Itanium processors can be included in internal benchmarking of actual enterprise workloads to determine which processor and platform are best.

Challenges

One ongoing challenge for Intel will be to clearly communicate which workloads are best supported by each of its product lines. That is, it is incumbent upon Intel to continue translating technical innovations and features into value propositions that influence IT planners. IDC agrees with Intel that sharp differentiation will not hold for all workloads or for all tiers in the enterprise server hierarchy. In midspectrum — for example, when provisioning servers for midtier enterprise workloads — IT planners will need to look beyond simple performance to decide which server, platform, and processor fit best. Some IT organizations will be of the mind-set that servers are a commodity with little or no differentiation.

IDC has forecast significant consolidation among manufacturers of processors and has argued that, at the system level, this consolidation would neither reduce consumers' choice nor increase price. Our forecasts remain in force. The message for consumers is that, at the system level, choices are expanding. The challenge for Intel is to compete as a component provider in a market that is increasingly competitive and filled with IA-32 and other processor offerings.

Intel will need to deliver on forecast performance and pricing for Xeon and Itanium processors. Proof points based on customer experience will be needed to support these best estimates as OEMs begin shipping new server products using the Xeon processor with 64-bit extensions.

Meeting the Challenges

Intel's OEM and software ecosystem is large and trusted in the IT industry, and members of this ecosystem have the ability to help make the Xeon processor with 64-bit extensions successful. Intel and industry partners continue to invest in enabling Itanium software, with over 1,000 applications now in production. Intel's ongoing investments in its developer community should pay off as the many players in the IT food chain turn their attention to this new offering.

Bringing the Xeon processor into the marketplace at this time is Intel's direct response to competing products and to customers who require extended address space and other performance enhancements for an IA-32 instruction set processor. Intel has demonstrated the ability to compete effectively in terms of price/performance, product quality, and platform focus.

Conclusion

Innovation in processor design at Intel is not limited to the Itanium processor family. Intel's new and enhanced Xeon processor will play an important role in the company's product lines, in the server products delivered by system suppliers, and in the datacenter where multitudes of IA-32 applications will move ahead to a faster and more flexible platform. The Intel Itanium processor remains the company's high-performance processor offering, and platforms using both Xeon and Itanium processors will be used across the spectrum of datacenter workloads, with Itanium processor–based servers taking on the most demanding workloads.

Copyright Notice